



REC'D 21 JUL 2003

WIPO

PCT

# BREVET D'INVENTION

## CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

### COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 20 MARS 2003

Pour le Directeur général de l'Institut  
national de la propriété industrielle  
Le Chef du Département des brevets

A handwritten signature in black ink, appearing to read 'M. Planche', enclosed within a large, loopy oval stroke.

Martine PLANCHE

#### DOCUMENT DE PRIORITÉ

PRÉSENTÉ OU TRANSMIS  
CONFORMÉMENT À LA  
RÈGLE 17.1.a) OU b)

INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

SIEGE  
26 bis, rue de Saint Petersburg  
75800 PARIS cedex 08  
Téléphone : 33 (0)1 53 04 53 04  
Télécopie : 33 (0)1 53 04 45 23  
www.inpi.fr

BEST AVAILABLE COPY

REMISE DES PIÈCES DATE <b>29 MARS 2002</b> LIEU <b>99</b> N° ENREGISTREMENT <b>0204285</b> NATIONAL ATTRIBUE PAR L'INPI DATE DE DÉPÔT ATTRIBUÉE PAR L'INPI <b>29 MARS 2002</b>	Réservé à l'INPI NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE A QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE  Monsieur MAILLET Alain Cabinet LE GUEN & MAILLET 5, place Newquay B.P. 70250 35802 DINARD CEDEX
--	---

Vos références pour ce dossier : 8263

(facultatif)

Confirmation d'un dépôt par télécopie	<input checked="" type="checkbox"/> N° attribué par l'INPI à la télécopie
<b>② NATURE DE LA DEMANDE</b>	<b>Cochez l'une des 4 cases suivantes</b>
Demande de brevet	<input checked="" type="checkbox"/>
Demande de certificat d'utilité	<input type="checkbox"/>
Demande divisionnaire	<input type="checkbox"/>
<i>Demande de brevet initiale</i> <i>On demande de certificat d'utilité</i> <i>initiale</i>	N° N° Date Date
Transformation d'une demande de Brevet européen <i>Demande de brevet initiale</i>	<input type="checkbox"/> N° Date

**③ TITRE DE L'INVENTION (200 caractères ou espaces maximum)**

Procédé de traduction de données autorisant une gestion de mémoire simplifiée

<b>④ DECLARATION DE PRIORITE OU REQUETE DU BENEFICE DE LA DATE DE DEPOT D'UNE DEMANDE ANTERIEURE FRANCAISE</b>	Pays ou organisation Date Pays ou organisation Date Pays ou organisation Date <input type="checkbox"/> s'il y a d'autres priorités, cochez la case et utilisez l'imprimé "Suite"
<b>⑤ DEMANDEUR</b>	<input type="checkbox"/> s'il y a d'autres demandeurs, cochez la case et utilisez l'imprimé "suite"
Nom ou dénomination social	FRANCE TELECOM
Prénoms	
Forme Juridique	S.A.
N° SIRET	380129866 RC Paris
Code APE-NAF	
Adresse	Rue 6, place d'Alleray Code postal et ville 75015 PARIS
Pays	FRANCE
Nationalité	Française
N° de téléphone (facultatif)	
N° de télécopie (facultatif)	
Adresse électronique (facultatif)	

**BREVET D'INVENTION  
CERTIFICAT D'UTILITÉ**

REQUÊTE EN DELIVRANCE 2/2

REMISE DES PIÈCES		DATE <b>29 MARS 2002</b> <span style="border: 1px solid black; padding: 2px;">Réserve à l'INPI</span>	
LIEU <b>99</b>		<b>0204285</b>	
N° ENREGISTREMENT NATIONAL ATTRIBUE PAR L'INPI		100 540W 260929	
<b>Vos références pour ce dossier :</b>		<b>8263</b>	
<i>(facultatif)</i>			
<b>⑥ MANDATAIRE</b>			
Nom		MAILLET	
Prénom		Alain	
Cabinet ou Société		Cabinet LE GUEN & MAILLET	
N° de pouvoir permanent et/ou de lien contractuel			
Adresse	Rue	5, place Newquay B.P. 70250	
	Code postal et ville	35802	DINARD Cedex
N° de téléphone <i>(facultatif)</i>		02 99 46 55 19	
N° de télécopie <i>(facultatif)</i>		02 99 46 41 80	
Adresse électronique <i>(facultatif)</i>		office@leguenmaillet.com	
<b>⑦ INVENTEUR (S)</b>			
Les inventeurs sont les demandeurs		<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non Dans ce cas fournir une désignation d'inventeur (s) séparée	
<b>⑧ RAPPORT DE RECHERCHE</b>		Uniquement pour une demande de brevet (y compris division et transformation)	
Etablissement immédiat ou établissement différé		<input checked="" type="checkbox"/> <input type="checkbox"/>	
Paiement échelonné de la redevance		Paiement en trois versements, uniquement pour les personnes physiques <input type="checkbox"/> Oui <input type="checkbox"/> Non	
<b>⑨ REDUCTION DU TAUX DES REDEVANCES</b>		Uniquement pour les personnes physiques. <input type="checkbox"/> Requête pour la première fois pour cette invention <i>(joindre un avis de non-impression)</i> <input type="checkbox"/> Requête antérieurement à ce dépôt <i>(joindre une copie de la décision d'admission pour cette invention ou indiquer la référence)</i>	
Si vous avez utilisé l'imprimé "suite", Indiquez le nombre de pages jointes			
<b>⑩ SIGNATURE DU DEMANDEUR OU DU MANDATAIRE</b> (Nom et qualité du signataire)		<b>VISA DE LA PREFECTURE OU DE L'INPI</b>	
MAILLET Alain 92 30360		L. GUICHET	

La présente invention concerne un procédé de traduction de données d'entrée en au moins une séquence lexicale de sortie, incluant une étape de décodage des données d'entrée au cours de laquelle des entités lexicales dont lesdites données sont représentatives sont identifiées au moyen d'au moins un modèle.

5 De tels procédés sont communément utilisés dans des applications de reconnaissance de parole, où au moins un modèle est mis en œuvre pour reconnaître des symboles acoustiques présents dans les données d'entrée, un symbole pouvant être constitué par exemple par un ensemble de vecteurs de paramètres d'un espace acoustique continu, ou encore par un label attribué à une entité sous-lexicale.

10 Dans certaines applications, le qualificatif "lexical" s'appliquera à une phrase considérée dans son ensemble, en tant que suite de mots, et les entités sous-lexicales seront alors des mots, alors que dans d'autres applications, le qualificatif "lexical" s'appliquera à un mot, et les entités sous-lexicales seront alors des phonèmes ou encore des syllabes aptes à former de tels mots, si ceux-ci sont de nature littérale, ou  
15 des chiffres, si les mots sont de nature numérique, c'est-à-dire des nombres.

Une première approche pour opérer une reconnaissance de parole consiste à utiliser un type particulier de modèle qui présente une topologie régulière et est destiné à apprendre toutes les variantes de prononciation de chaque entité lexicale, c'est-à-dire par exemple un mot, inclus dans le modèle. Selon cette première  
20 approche, les paramètres d'un ensemble de vecteurs acoustiques propre à chaque symbole d'entrée correspondant à un mot inconnu doivent être comparés à des ensembles de paramètres acoustiques correspondant chacun à l'un des très nombreux symboles contenus dans le modèle, afin d'identifier un symbole modélisé auquel correspond le plus vraisemblablement le symbole d'entrée. Une telle approche garantit  
25 en théorie un fort taux de reconnaissance si le modèle utilisé est bien conçu, c'est-à-dire quasi-exhaustif, mais une telle quasi-exhaustivité ne peut être obtenue qu'au prix d'un long processus d'apprentissage du modèle, qui doit assimiler une énorme quantité de données représentatives de toutes les variantes de prononciation de chacun des mots inclus dans ce modèle. Cet apprentissage est en principe réalisé en faisant  
30 prononcer par un grand nombre de personnes tous les mots d'un vocabulaire donné, et

à enregistrer toutes les variantes de prononciation de ces mots. Il apparaît clairement que la construction d'un modèle lexical quasi-exhaustif n'est pas envisageable en pratique pour des vocabulaires présentant une taille supérieure à quelques centaines de mots.

5 Une deuxième approche a été conçue dans le but de réduire le temps d'apprentissage nécessaire aux applications de reconnaissance de parole, réduction qui est essentielle à des applications de traduction sur de très grands vocabulaires pouvant contenir plusieurs centaines de milliers de mots, laquelle deuxième approche consiste à opérer une factorisation des entités lexicales en les considérant comme des  
10 assemblages d'entités sous-lexicales, à générer un modèle sous-lexical modélisant lesdites entités sous-lexicales en vue de permettre leur identification dans les données d'entrée, et un modèle d'articulation modélisant différentes combinaisons possibles de ces entités sous-lexicales. Selon cette deuxième approche, un nouveau modèle dynamique formant le modèle d'articulation est constitué à partir de chaque entité  
15 sous-lexicale nouvellement identifiée dans les données d'entrée, lequel modèle dynamique rend compte de tous les assemblages rendus possibles en partant de l'entité sous-lexicale considérée, et détermine une valeur de vraisemblance pour chaque assemblage possible.

Une telle approche, décrite par exemple au chapitre 16 du manuel "Automatic  
20 Speech and Speaker Recognition" édité par Kluwer Academic Publishers, permet de réduire considérablement, par rapport au modèle utilisé dans le cadre de la première approche décrite plus haut, les durées individuelles des processus d'apprentissage du modèle sous-lexical et du modèle d'articulation, car chacun de ces modèles présente une structure simple par rapport au modèle lexical utilisé dans la première approche.

25 Cependant, dans la plupart des implémentations connues de la deuxième approche décrite ci-dessus, le modèle sous-lexical est dupliqué à de multiples reprises dans le modèle d'articulation. Ceci peut être aisément compris en considérant un exemple où l'unité lexicale est une phrase et les unités sous-lexicales sont des mots. Si le modèle d'articulation est d'un type bi-gramme, c'est-à-dire qu'il rend compte de  
30 possibilités d'assemblage de deux mots successifs et de probabilités d'existence de

tels assemblages, chaque mot retenu à l'issue de la sous-étape d'identification devra être étudié, en référence au modèle d'articulation, avec tous les autres mots retenus ayant pu précéder le mot considéré. Si  $P$  mots ont été retenus à l'issue de la sous-étape d'identification,  $P$  couples de mots devront être construits pour chaque mot à identifier, avec  $P$  valeurs de probabilité d'existence, chacune associée à un couple possible. Dans le cas d'un modèle d'articulation plus réaliste de type tri-gramme, qui rend compte de possibilités d'assemblage de trois mots successifs et de probabilités d'existence de tels assemblages, le modèle d'articulation devra comporter, pour chaque mot à identifier,  $P$  fois  $P$  triplets de mots avec autant de valeurs de probabilité d'existence. Les modèles d'articulation mis en œuvre dans la deuxième approche ont donc une structure simple, mais représentent un volume considérable de données à mémoriser, à mettre à jour et à consulter. On conçoit aisément que la création et l'exploitation de tels modèles donne lieu à des accès mémoire dont la gestion est rendue complexe par le volume de données à traiter, et par la répartition desdites données. Dans des applications de type langage naturel, pour lesquelles des modèles plus réalistes de type  $N$ -gramme, où  $N$  est le plus souvent supérieur à deux, sont mis en œuvre, les accès mémoire évoqués précédemment présentent des temps d'exécution incompatibles avec des contraintes de type "temps réel" nécessitant des accès mémoire très rapides.

Par ailleurs, chaque mot peut lui-même être considéré vis-à-vis de syllabes ou de phonèmes qui le composent comme une entité lexicale d'un niveau inférieur à celui d'une phrase, entité lexicale pour la modélisation de laquelle il faut également recourir à un modèle d'articulation de type  $N$ -gramme avec plusieurs dizaines d'entités sous-lexicales possibles dans le cas des phonèmes.

Il apparaît clairement que les multiples duplications des modèles sous-lexicaux auxquelles font appel les modèles d'articulation dans les implémentations connues de la deuxième approche prohibent l'utilisation de celle-ci dans des applications de reconnaissance de parole dans le cadre d'applications de type très grands vocabulaires, qui comportent plusieurs centaines de milliers de mots.

L'invention a pour but de remédier dans une large mesure à cet inconvénient, en proposant un procédé de traduction qui ne nécessite pas de multiples duplications de modèles sous-lexicaux pour valider des assemblages d'entités sous-lexicales, et simplifie ainsi l'implémentation dudit procédé de traduction, et en particulier la  
 5 gestion d'accès mémoire utiles à ce procédé.

En effet, un procédé de traduction conforme au paragraphe introductif, incluant une étape de décodage au cours de laquelle des entités sous-lexicales dont les données d'entrée sont représentatives sont identifiées au moyen d'un premier modèle construit sur la base d'entités sous-lexicales prédéterminées, et au cours de laquelle sont  
 10 générées, au fur et à mesure que les entités sous-lexicales sont identifiées et en référence à au moins un deuxième modèle construit sur la base d'entités lexicales, diverses combinaisons possibles desdites entités sous-lexicales, est caractérisé selon l'invention en ce que l'étape de décodage inclut une sous-étape de mémorisation d'une pluralité de combinaisons possibles desdites entités sous-lexicales, la  
 15 combinaison la plus vraisemblable étant destinée à former la séquence lexicale de sortie.

Du fait que divers assemblages d'entités sous-lexicales sont mémorisés au fur et à mesure que ces entités sont produites, il n'est plus nécessaire de construire après identification de chacune desdites entités sous-lexicales un modèle dynamique  
 20 reprenant toutes les entités sous-lexicales possibles, ce qui permet d'éviter les duplications évoquées plus haut et les problèmes de gestion mémoire y afférant.

La possibilité de mémoriser plusieurs combinaisons différentes permet de garder une trace de plusieurs assemblages possibles d'entités sous-lexicales, chacun présentant une vraisemblance propre à l'instant où cet assemblage est généré, laquelle  
 25 vraisemblance pouvant être affectée favorablement ou défavorablement après analyse de sous-entités lexicales ultérieurement produites. Ainsi, une sélection d'un assemblage présentant la plus forte vraisemblance à un instant donné, mais qui sera finalement jugé peu vraisemblable à la lumière d'entités sous-lexicales ultérieures ne provoquera pas une élimination systématique d'autres assemblages, qui pourront  
 30 finalement s'avérer plus pertinents. Cette variante de l'invention permet donc de

conserver des données représentant, sous forme de différents historiques, différentes interprétations des données d'entrée, interprétations dont la plus vraisemblable pourra être identifiée et retenue pour former la séquence lexicale de sortie lorsque toutes les entités sous-lexicales auront elles-même été identifiées.

- 5 Dans un mode de réalisation particulier de cette variante de l'invention, la mémorisation d'une combinaison est assujettie à une validation opérée en référence au moins au deuxième modèle.

Ce mode de réalisation permet de réaliser de manière simple un filtrage des assemblages qui paraissent peu vraisemblables à la lumière du deuxième modèle.

- 10 Seuls seront retenus et mémorisés les assemblages les plus plausibles, les autres assemblages n'étant pas mémorisés et donc pas ultérieurement pris en considération.

- Dans une variante de ce mode de réalisation, la validation de mémorisation pourra être effectuée en référence à plusieurs modèles de niveaux équivalents et/ou différentes, un niveau rendant compte de la nature sous-lexicale, lexicale ou encore  
15 grammaticale d'un modèle.

Dans un mode de réalisation particulièrement avantageux de cette variante de l'invention, une validation de mémorisation d'une combinaison est accompagnée d'une attribution à la combinaison à mémoriser d'une valeur de probabilité représentative de la vraisemblance de ladite combinaison.

- 20 Ce mode de réalisation permet de moduler la nature binaire du filtrage opérée par la validation ou l'absence de validation de la mémorisation d'une combinaison, en affectant une appréciation quantitative à chaque combinaison mémorisée. Ceci permettra une meilleure appréciation de la vraisemblance des diverses combinaisons qui auront été mémorisées, et donc une traduction de meilleure qualité des données  
25 d'entrée.

On pourra en outre prévoir que différentes opérations de validation portant sur différentes combinaisons relatives à un même état du premier modèle sont exécutées de façon contiguë dans le temps.



Ceci permettra de réduire encore le volume des accès mémoire et des duplications de calcul, en traitant en une seule fois toute une famille d'informations qu'il faudra sinon mémoriser et lire à de multiples reprises.

5 Dans un mode de réalisation particulier de l'invention, l'étape de décodage met en œuvre un algorithme de Viterbi appliqué à un premier modèle de Markov constitué d'entités sous-lexicales, sous contrôle dynamique d'un deuxième modèle de Markov représentatif de combinaisons possibles d'entités sous-lexicales.

10 Ce mode de réalisation est avantageux en ce qu'il utilise des moyens éprouvés et individuellement connus de l'homme du métier, le contrôle dynamique obtenu grâce au deuxième modèle de Markov permettant de valider les assemblages d'entités sous-lexicales au fur et à mesure que lesdites entités sont identifiées au moyen de l'algorithme de Viterbi, ce qui évite d'avoir à construire après identification de chaque entité sous-lexicale un nouveau modèle dynamique reprenant toutes les entités sous-lexicales possibles semblable à ceux utilisés dans les implémentations connues de la  
15 deuxième approche évoquée plus haut.

L'invention concerne également un système de reconnaissance de signaux acoustiques mettant en œuvre un procédé tel que décrit ci-dessus.

20 Les caractéristiques de l'invention mentionnées ci-dessus, ainsi que d'autres, apparaîtront plus clairement à la lecture de la description suivante d'un exemple de réalisation, ladite description étant faite en relation avec les dessins joints, parmi lesquels :

La Fig.1 est un schéma fonctionnel décrivant un système de reconnaissance acoustique dans lequel un procédé conforme à l'invention est mis en œuvre,

25 La Fig.2 est un schéma fonctionnel décrivant un décodeur destiné à exécuter une première étape de décodage dans ce mode de mise en œuvre particulier de l'invention, et

La Fig.3 est un schéma fonctionnel décrivant un décodeur destiné à exécuter une deuxième étape de décodage conforme au procédé selon l'invention.

30 La Fig.1 représente schématiquement un système SYST de reconnaissance acoustique selon un mode de mise en œuvre particulier de l'invention, destiné à

traduire un signal acoustique d'entrée ASin en une séquence lexicale de sortie OUTSQ. Le signal d'entrée ASin est constitué par un signal électronique analogique, qui pourra provenir par exemple d'un microphone non représenté sur la figure. Dans le mode de réalisation décrit ici, le système SYST inclut un étage d'entrée FE, contenant un dispositif de conversion analogique/numérique ADC, destiné à fournir un signal numérique ASin(1:n), formé d'échantillons ASin(1), ASin(2)...ASin(n) codés chacun sur b bits, et représentatif du signal acoustique d'entrée ASin, et un module d'échantillonnage SA, destiné à convertir le signal acoustique numérisé ASin(1:n) en une séquence de vecteurs acoustiques AVin, chaque vecteur étant muni de composantes AV1, AV2...AVr où r est la dimension d'un espace acoustique défini pour une application donnée à laquelle le système de traduction SYST est destiné, chacune des composantes AVi (pour i=1 à r) étant évaluée en fonction de caractéristiques propres à cet espace acoustique.

Le système SYST inclut en outre un premier décodeur DEC1, destiné à fournir une sélection Int1, Int2...IntK d'interprétations possibles de la séquence de vecteurs acoustiques AVin en référence à un modèle MD1 construit sur la base d'entités sous-lexicales prédéterminées.

Le système SYST inclut de plus un deuxième décodeur DEC2 dans lequel un procédé de traduction conforme à l'invention est mis en œuvre en vue d'analyser des données d'entrée constituées par les vecteurs acoustiques AVin en référence à un premier modèle construit sur la base d'entités sous-lexicales prédéterminées, par exemple le modèle MD1, et en référence à au moins un deuxième modèle MD2 construit sur la base d'entités lexicales représentatives des interprétations Int1, Int2...IntK sélectionnées par le premier décodeur DEC1, en vue d'identifier celle desdites interprétations qui devra constituer la séquence lexicale de sortie OUTSQ.

La fig.2 représente plus en détail le premier décodeur DEC1, qui inclut une première machine de Viterbi VM1, destinée à exécuter une première sous-étape de décodage de la séquence de vecteurs acoustiques AVin représentative du signal acoustique d'entrée et préalablement générée par l'étage d'entrée FE, laquelle séquence sera en outre avantageusement mémorisée dans une unité de stockage

MEM1 pour des raisons qui apparaîtront dans la suite de l'exposé. La première sous-étape de décodage est opérée en référence à un modèle de Markov MD11 autorisant en boucle toutes les entités sous-lexicales, de préférence tous les phonèmes de la langue dans laquelle le signal acoustique d'entrée doit être traduit si l'on considère que  
 5 les entités lexicales sont des mots, les entités sous-lexicales étant représentées sous forme de vecteurs acoustiques prédéterminés.

La première machine de Viterbi VM1 est apte à restituer une séquence de phonèmes Phsq qui constitue la plus proche traduction phonétique de la séquence de vecteurs acoustiques AVin. Les traitements ultérieurs réalisés par le premier décodeur  
 10 DEC1 se feront ainsi au niveau phonétique, et non plus au niveau vectoriel, ce qui réduit considérablement la complexité desdits traitements, chaque vecteur étant une entité multidimensionnelle présentant  $r$  composantes, tandis qu'un phonème peut en principe être identifié par un label unidimensionnel qui lui est propre, comme par exemple un label "OU" attribué à une voyelle orale "u", ou un label "CH" attribué à  
 15 une consonne frictive non-voisée "f". La séquence de phonèmes Phsq générée par la première machine de Viterbi VM1 est ainsi constituée d'une succession de labels plus aisément manipulables que ne le seraient des vecteurs acoustiques.

Le premier décodeur DEC1 inclut une deuxième machine de Viterbi VM2 destinée à exécuter une deuxième sous-étape de décodage de la séquence de phonèmes  
 20 Phsq générée par la première machine de Viterbi VM1. Cette deuxième étape de décodage est opérée en référence à un modèle de Markov MD12 constitué de transcriptions sous-lexicales d'entités lexicales, c'est-à-dire dans cet exemple de transcriptions phonétiques de mots présents dans le vocabulaire de la langue dans laquelle le signal acoustique d'entrée doit être traduit. La deuxième machine de  
 25 Viterbi est destinée à interpréter la séquence de phonèmes Phsq, qui est fortement bruitée du fait que le modèle MD11 utilisé par la première machine de Viterbi VM1 est d'une grande simplicité, et met en œuvre des prédictions et des comparaisons entre des suites de labels de phonèmes contenus dans la séquence de phonèmes Phsq et diverses combinaisons possibles de labels de phonèmes prévues dans le modèle de  
 30 Markov MD12. Bien qu'une machine de Viterbi ne restitue usuellement que celle des

séquences qui présente la plus grande probabilité, la deuxième machine de Viterbi VM2 mise en œuvre ici restituera avantageusement toutes les séquences de phonèmes lsq1, lsq2...lsqN que ladite deuxième machine VM2 aura pu reconstituer, avec des valeurs de probabilité associées p1, p2...pN qui auront été calculées pour lesdites  
 5 séquences et seront représentatives de la fiabilité des interprétations du signal acoustique que ces séquences représentent.

Toutes les interprétations possibles lsq1, lsq2...lsqN étant rendues automatiquement disponibles à l'issue de la deuxième sous-étape de décodage, une sélection de K interprétations Int1, Int2...IntK qui présentent les plus fortes valeurs de  
 10 probabilité est aisée quelle que soit la valeur de K qui aura été choisie.

Les première et deuxième machines de Viterbi VM1 et VM2 peuvent fonctionner en parallèle, la première machine de Viterbi VM1 générant alors au fur et à mesure des labels de phonèmes qui seront immédiatement pris en compte par la deuxième machine de Viterbi VM2, ce qui permet de réduire le délai total perçu par  
 15 un utilisateur du système nécessaire à la combinaison des première et deuxième sous-étapes de décodage en autorisant la mise en œuvre de l'ensemble des ressources de calcul nécessaires au fonctionnement du premier décodeur DEC1 dès que les vecteurs acoustiques AVin représentatifs du signal acoustique d'entrée apparaissent, et non pas après qu'ils aient été entièrement traduits en une séquence complète de phonèmes  
 20 Phsq par la première machine de Viterbi VM1.

La Fig.3 représente plus en détail un deuxième décodeur DEC2 conforme à un mode de réalisation particulier de l'invention. Ce deuxième décodeur DEC2 inclut une troisième machine de Viterbi VM3 destinée à analyser la séquence de vecteurs acoustiques AVin représentative du signal acoustique d'entrée préalablement  
 25 mémorisée dans l'unité de stockage MEM1.

A cet effet, la troisième machine de Viterbi VM3 est destinée à exécuter une sous-étape d'identification au cours de laquelle les entités sous-lexicales dont les vecteurs acoustiques AVin sont représentatifs sont identifiées au moyen d'un premier modèle construit sur la base d'entités sous-lexicales prédéterminées, dans cet exemple

le modèle de Markov MD11 mis en œuvre dans le premier décodeur et déjà décrit plus haut.

La troisième machine de Viterbi VM3 génère en outre, au fur et à mesure que ces entités sont identifiées et en référence à au moins un modèle de Markov spécifique MD3 construit sur la base d'entités lexicales, diverses combinaisons possibles des entités sous-lexicales, la combinaison la plus vraisemblable étant destinée à former la séquence lexicale de sortie OUTSQ. Le modèle de Markov spécifique MD3 est ici spécialement généré à cet effet par un module de création de modèle MGEN, et est uniquement représentatif d'assemblages possibles de phonèmes au sein des séquences de mots formées par les diverses interprétations phonétiques Int1, Int2,...IntK du signal acoustique d'entrée délivrées par le premier décodeur, lesquels assemblages sont représentés par des sous-modèles extraits du modèle lexical MD2 par le module de création de modèle MGEN. Le modèle de Markov spécifique MD3 présente donc une taille restreinte du fait de sa spécificité.

Lorsque la troisième machine de Viterbi VM3 se trouve dans un état ni donné, auquel sont associés un historique hp et une valeur de probabilité Sp, s'il existe dans le modèle de Markov MD11 une transition dudit état ni vers un état nj munie d'un marqueur M, lequel marqueur pouvant par exemple être constitué par le label d'un phonème dont le dernier état est ni ou d'un phonème dont le premier état est nj, la troisième machine de Viterbi VM3 associera à l'état nj un nouvel historique hq et une nouvelle valeur de probabilité Sq qui seront générés en référence au modèle spécifique MD3, sur la base de l'historique hp, de sa valeur de probabilité associée Sp et du marqueur M, la valeur de probabilité Sp pouvant en outre être également modifiée en référence au modèle de Markov MD11. Cette opération sera répétée pour tous les historiques associés à l'état ni. Si un même historique hk est associé à plusieurs reprises à un même état du modèle de Markov MD11 avec différentes valeurs de probabilité Sp1,...Spq, conformément à l'algorithme de Viterbi, seule la valeur de probabilité la plus élevée sera conservée et attribuée en tant que valeur de probabilité Sp à l'historique hk.

Chaque état  $n_j$  est mémorisé dans une unité de stockage MEM2 avec ses différents historiques  $h_q$  et une valeur de probabilité  $S_q$  propre à chaque historique, et ce jusqu'à ce que la troisième machine de Viterbi VM3 ait identifié tous les phonèmes contenus dans la séquence de vecteurs acoustiques d'entrée AVin et ait atteint un  
 5 dernier état  $n_f$  au fil d'une pluralité d'historiques  $h_f$  représentant les diverses combinaisons possibles des phonèmes identifiés. Celui de ces historiques auquel aura été attribuée la plus forte valeur de probabilité  $S_{f_{\max}}$  sera retenu par un décodeur de mémoire MDEC pour former la séquence lexicale de sortie OUTSQ.

Le modèle de Markov MD3 opère donc un contrôle dynamique permettant de  
 10 valider les assemblages de phonèmes au fur et à mesure que lesdits phonèmes sont identifiés par la troisième machine de Viterbi VM3, ce qui évite d'avoir à dupliquer ces phonèmes pour former des modèles tels ceux utilisés dans les implémentations connues de la deuxième approche évoquée plus haut. De la sorte, les accès aux unités de stockage MEM1 et MEM2, ainsi qu'au différents modèles de Markov MD11,  
 15 MD12, MD2 et MD3 mis en œuvre dans l'exemple décrit ci-dessus nécessitent une gestion peu complexe, du fait de la simplicité de structure desdits modèles et des informations destinées à être mémorisées et lues dans lesdites unités de stockage. Ces accès mémoire peuvent donc être exécutés suffisamment rapidement pour rendre le système décrit dans cet exemple apte à accomplir des traductions en temps réel de  
 20 données acoustiques d'entrée en séquences lexicales de sortie.

Bien que l'invention ait été décrite ici dans le cadre d'une application au sein d'un système incluant deux décodeurs disposés en cascade, il est tout-à-fait envisageable, dans d'autres modes de mise en œuvre de l'invention, de n'utiliser qu'un unique décodeur semblable au deuxième décodeur décrit plus haut, qui pourra  
 25 par exemple opérer une analyse acoustico-phonétique et mémoriser, au fur et à mesure que des phonèmes seront identifiés, diverses combinaisons possibles desdits phonèmes, la combinaison de phonèmes la plus vraisemblable étant destinée à former la séquence lexicale de sortie.

## REVENDICATIONS

1) Procédé de traduction de données d'entrée en au moins une séquence lexicale de sortie, incluant une étape de décodage des données d'entrée au cours de laquelle des entités sous-lexicales dont lesdites données sont représentatives sont identifiées au moyen d'un premier modèle construit sur la base d'entités sous-lexicales prédéterminées, et au cours de laquelle sont générées, au fur et à mesure que les entités sous-lexicales sont identifiées et en référence à au moins un deuxième modèle construit sur la base d'entités lexicales, diverses combinaisons possibles desdites entités sous-lexicales,

procédé caractérisé en ce que l'étape de décodage inclut une sous-étape de mémorisation d'une pluralité de combinaisons possibles desdites entités sous-lexicales, la combinaison la plus vraisemblable étant destinée à former la séquence lexicale de sortie.

2) Procédé de traduction selon la revendication 1, caractérisé en ce que la mémorisation d'une combinaison est assujettie à une validation opérée en référence au moins au deuxième modèle.

3) Procédé de traduction selon la revendication 2, caractérisé en ce qu'une validation de mémorisation d'une combinaison est accompagnée d'une attribution à la combinaison à mémoriser d'une valeur de probabilité représentative de la vraisemblance de ladite combinaison.

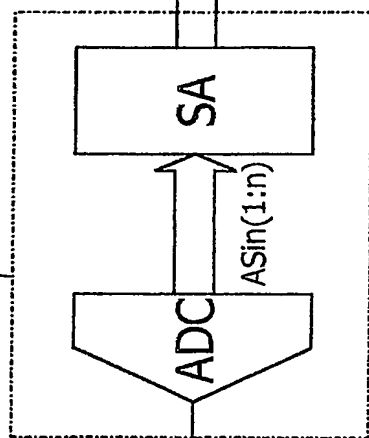
4) Procédé de traduction selon l'une des revendications 2 ou 3, caractérisé en ce que différentes opérations de validation portant sur différentes combinaisons relatives à un même état du premier modèle sont exécutées de façon contiguë dans le temps.

5) Procédé de traduction selon la revendication 1, caractérisé en ce que l'étape de décodage met en œuvre un algorithme de Viterbi appliqué à un premier modèle de Markov constitué d'entités sous-lexicales, sous contrôle dynamique d'un deuxième modèle de Markov représentatif de combinaisons possibles d'entités sous-lexicales.

6) Système de reconnaissance vocale mettant en œuvre un procédé de traduction conforme à l'une des revendications 1 à 5.

SYST

FE



AVin

$\begin{bmatrix} AV1 \\ AV2 \\ \vdots \\ AVr \end{bmatrix}$

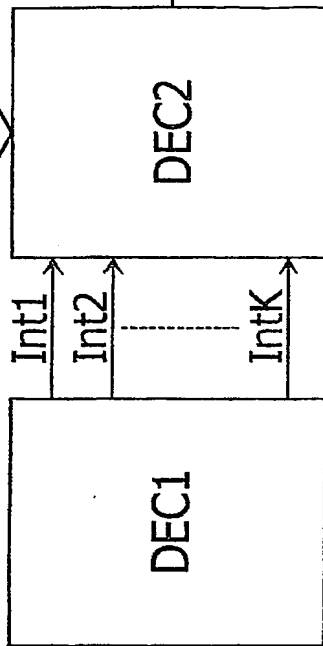
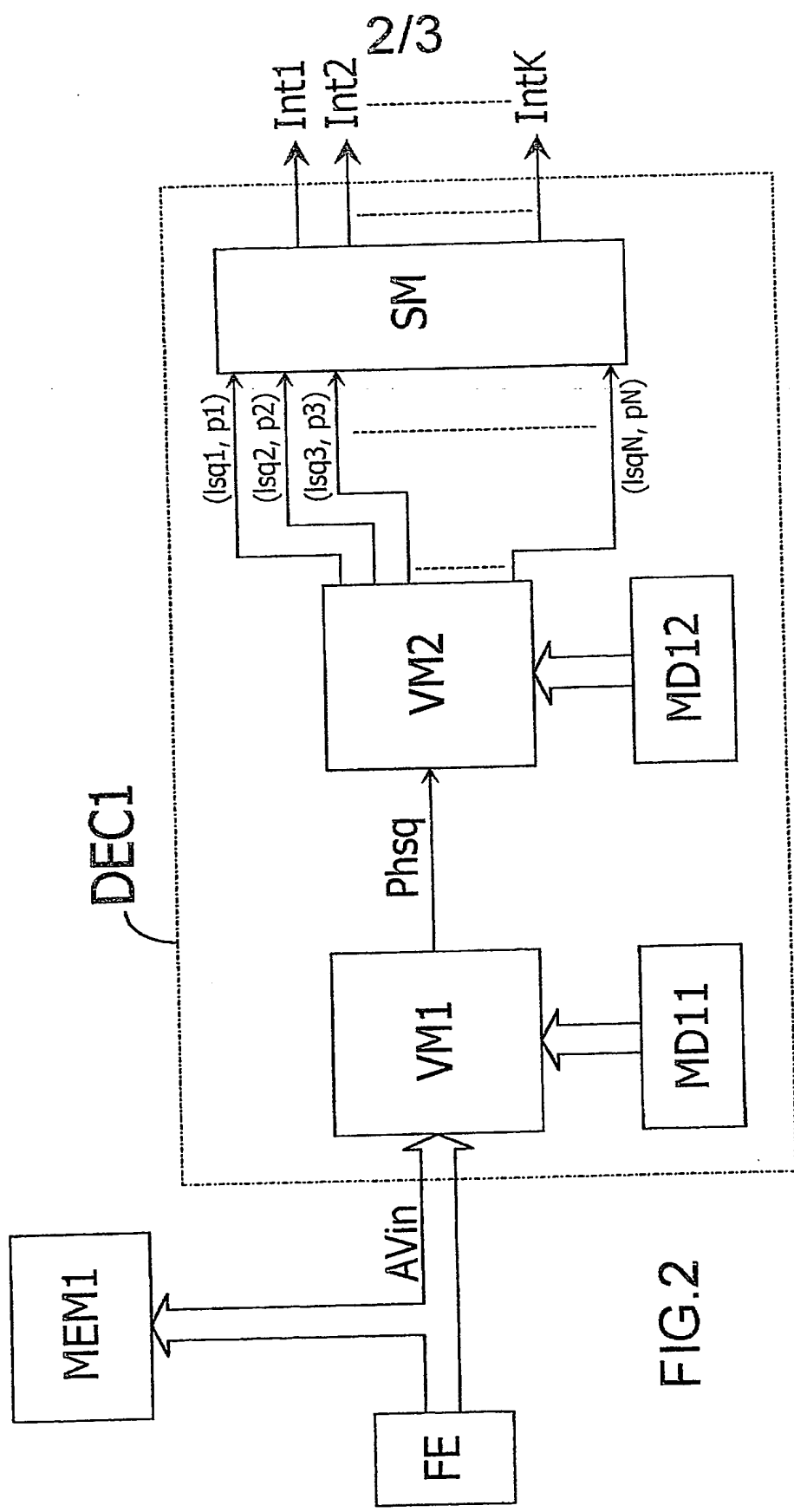


FIG.1





2/3

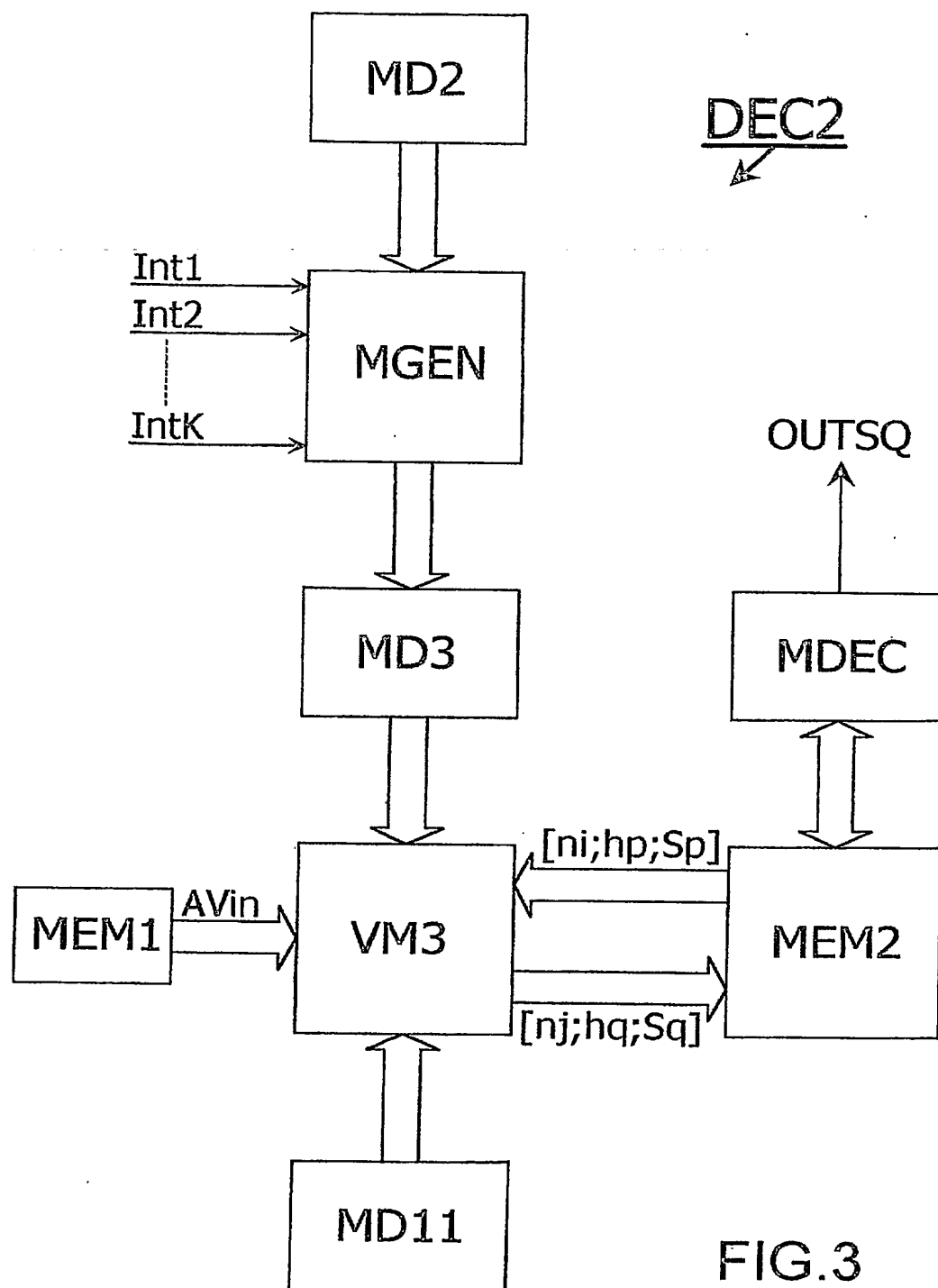


FIG.3

reçue le 24/04/02

**BREVET D'INVENTION  
CERTIFICAT D'UTILITE**

Code de la propriété intellectuelle – Livre VI

**INPI**

INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

DEPARTEMENT DES BREVETS

26 bis, rue de Saint Pétersbourg

75800 Paris Cedex 08

Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54



N°11235-02

**DESIGNATION DE L'INVENTEUR (S)** Page N° 1/1

(si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

LD 113 W 2000

Vos références pour ce dossier (facultatif)		8263	
N° D'ENREGISTREMENT NATIONAL 02 04285			
TITRE DE L'INVENTION (200 caractères ou espaces maximum)  Procédé de traduction de données autorisant une gestion de mémoire simplifiée			
LE(S) DEMANDEUR(S) : FRANCE TELECOM S.A. 6, place d'Alleray 75015 PARIS			
DESIGNE (NT) EN TANT QU'INVENTEUR(S) : (Indiquez en haut à droite "page N°1/1" S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages).			
Nom		FERRIEUX	
Prénoms		Alexandre	
Adresse	Rue	4, Hent Al Lann	
	Code postal et ville	22560	PLEUMEUR BODOU
Société d'appartenance (facultatif)			
Nom		DELPHIN-POULAT	
Prénoms		Lionel	
Adresse	Rue	Résidence Kergemar A2	
	Code postal et ville	22300	LANNION
Société d'appartenance (facultatif)			
Nom			
Prénoms			
Adresse	Rue		
	Code postal et ville		
Société d'appartenance (facultatif)			
SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire)			
MAILLET Alain 92 30360			

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**